DOCUMENT RESUME

ED 456 665 FL 026 891

AUTHOR Watt, David L. E.; Lake, Deidre M.

TITLE Canadian Language Benchmarks-TOEFL Research Project: A Comparison Study of the Canadian Language Benchmarks Assessment and the Test of English as a Foreign Language.

INSTITUTION Calgary Univ. (Alberta). Faculty of Education.
SPONS AGENCY Alberta Learning, Edmonton.
PUB DATE 2000-00-00
NOTE 27p.; Funded through Alberta Learning, Language Training Programs.
PUB TYPE Reports - Research (143)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Admission Criteria; Adults; *Benchmarking; College Admission; *College Entrance Examinations; *Communicative Competence (Languages); Comparative Analysis; *English (Second Language); Foreign Countries; Second Language Instruction; Second Language Learning; Student Evaluation; *Test Validity
IDENTIFIERS Canada; *Test of English as a Foreign Language

ABSTRACT

This study is an examination of the test results of 90 academically oriented adult participants on the Test of English as a Foreign Language (TOEFL) and the Canadian Language Benchmarks Assessment (CBLA), to determine the comparability of performance on the two tests and the possibility of using two tests in the academic admissions process for colleges and universities. It is concluded that the two tests are measuring similar language constructs, but that each also adds something unique to the picture of the participants' English language proficiency. It was also found that length of residence in Canada had a significant predictive effect only in the listening/speaking section of the CBLA. This suggests that the CBLA is especially good for measuring the communicative competence of academically-oriented English language learners. Due to this finding, grounded in a sufficiently large, representative, and significant study group, it is now possible to make an empirical case to argue for the inclusion of the CBLA test in college and university admissions processes in Canada in order to give the fullest and fairest consideration of English language learners in the college and university applicant pool. (KFT)

Reproductions supplied by EDRS are the best that can be made from the original document.

ED 456 665

Canadian Language Benchmarks – TOEFL Research Project

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
D. Watt
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

A comparison study of the Canadian Language Benchmarks Assessment and the Test of English as a Foreign Language

2000

This project was funded through Alberta Learning, Language Training Programs

BEST COPY AVAILABLE

FL 026 891

ACKNOWLEDGEMENTS

We would like to acknowledge the following institutions, agencies and individuals for their invaluable support in this project:

- University of Calgary
  Registrar's Office
  Faculty of Engineering
  Faculty of Communication and Culture
  Faculty of Science
  English for Academics Purposes Program
- Prometric – Sylvan Learning Centre – TOEFL Testing Centre
- Mount Royal College Languages Institute
- Bow Valley College – TOEFL Preparation Program
- Southern Alberta Institute of Technology – International Academic Upgrading Program
- Maple Leaf Academy – TOEFL Preparation Program
- Rocky Mountain English Centre – TOEFL Preparation Program
- Canadian Conversation Institute – TOEFL Preparation Program
- Centre for Language Training and Assessment - Peele Board of Education
- Educational Testing Service – TOEFL Division
- Martek Assessments Limited

This project was funded through Alberta Learning, Language Training Programs

TABLE OF CONTENTS

Summary of the CLBA-TOEFL Research Report _____ 1-2

Overview of the Canadian Language Benchmarks Assessment _____ 3-5

Overview of the Test of English as a Second Language _____ 6-7

Description of the participants _____ 8-9

TOEFL score comparison for the participant sample _____ 10

CLBA benchmarks distribution for the sample population _____ 11-14

Correlations among CLBA and TOEFL scores _____ 15-17

Equating TOEFL performances with CLBA performance _____ 18-19

Technical note on statistical equating _____ 20

Conclusions and recommendations _____ 21-22

This project was funded through Alberta Learning, Language Training Programs

Background

The CLBA-TOEFL Research Project conducted an examination of the test results of 90 academically oriented adult participants on the Test of English as a Foreign Language (TOEFL) and the Canadian Language Benchmarks Assessment (CLBA), in order to determine the comparability of performances on the two tests and the possibility of using both tests in the academic admission process for colleges and universities. A sample of convenience was collected from voluntary participants who were either already admitted to universities and colleges, or who were seeking admission. Participants were requested to provide a recent TOEFL record sheet, to take the CLBA and to provide personal, educational and occupational background information. Each participant was assessed on the CLBA by one of the two researchers in the project. Both researchers hold national certification as CLBA assessors. Random double marking of test sections and unobtrusive interview observations were introduced as a means of increasing the inter-rater reliability.

The data were analyzed using SPSS (Statistical Procedures for the Social Sciences) in order to determine the following:

1. The descriptive profile of the participant sample, to determine applicability of the findings to a larger audience of academically oriented TOEFL takers.
2. The descriptive profile of performances on each of the two tests.
3. The correlations between the two tests and the relevant sections of the two tests, in order to establish the degree to which they are related and the unique role that each test plays in the assessment of English Language Proficiency (ELP).
4. The role that Length of Residence (LOR) has on the communicative ability as measured by each of the two tests, in order to gain insights that may be relevant to the ELP assessment of Landed Immigrants, New Canadians or other long term residents.
5. A statistical means of interpolating test scores on the two tests using a scaled equation method, to produce a heuristic concordance table for the TOEFL and CLBA.

Primary Findings

1. The participant sample proved to be highly representative of a general audience of academically oriented individuals who require TOEFL as an admission criterion to university and college programs. It represented 26 countries, 21 languages and an array of professional and educational experiences. Participant performance on the TOEFL was a close match to the inter-correlations of performance that have been established for the TOEFL by Educational Testing Service (ETS).
2. Stage II of the CLBA (Benchmarks 5-8) was able to measure and discriminate the ELP of between 75% and 92% of all the participants. 89% of all the individual section scores on the CLBA fell between the Benchmark 5-8 range. The remaining 11% of the performances were distinguishable as beyond the Benchmark 8 threshold and were identified as 8+ performance.

3. Across the wide range of TOEFL and CLBA scores (137-280, computer-based / 457-650, paper-based), there was evidence of a moderately strong correlation. The strength of the correlation suggests that the two tests are measuring similar language constructs, but that each also adds some unique information to a participant's profile of ELP.
4. A simple regression analysis, based on Length of Residence (LOR) and the various sections of the TOEFL and CLBA, found that LOR had a significant predictive effect on only the CLBA listening/speaking section. The uniqueness of this finding suggests a strong role for the CLBA in determining the communicative ability of academically oriented second language speakers of English, and may be an important consideration in determining the quality of admission equivalences that base themselves on LOR or resident study.
5. A heuristic concordance table for TOEFL and CLBA score comparison was statistically feasible, based on interpolation of the data. A test equation method used by ETS was applied to the data. The result is a table that offers a scaled comparison of TOEFL and CLBA scores. The heuristic concordance table provides a detailed foundation for discussions about the future use of the CLBA in college and university admission procedures.

Immediate Significance

This study is the first mid-sized CLBA-TOEFL comparison study with a sufficiently large participant base to generate both descriptive and inferential statistics regarding comparative test performance. It makes a significant contribution to the social policy goals of providing equitable access to university and college education for both international and immigrant second language applicants. The study provides baseline information in support of adding the nationally developed CLBA to the list of TOEFL equivalencies for admission purposes. It also paves the way for a similar examination and discussion of ELP standards that are presently used for accreditation by professional associations.

Recommendations

1. Undertake a broader scaled study to verify the heuristic table of concordance between the TOEFL and the CLBA.
2. Promote the use of the CLBA for admission purposes in Canadian universities, colleges and professional associations.
3. Undertake similar studies with professional associations in order to update English language proficiency standards for professional standing.
4. Compare the CLBA to other accepted equivalencies for university and college admission (e.g. three years of full time study in a Canadian institution, or five years of residency).
5. Include the findings of this research project in the formative development of the CLBA Stage III (Benchmarks 9-12).

Overview of the CLBA

The Canadian Language Benchmarks Assessment is designed to assess the English language proficiency levels of newcomers to Canada. The assessment is a task-based assessment for the placement of adult ESL learners into appropriate ESL programs and/or to determine their professional or academic readiness. The CLBA has three components:

- Listening/Speaking
- Reading
- Writing

The **CLBA Listening/Speaking Assessment** is designed so participants attempt a range of tasks of different types. This allows participants to demonstrate their proficiency and gives assessors sufficient evidence on which to base decisions. There are two stages; the first stage focuses on the participants' fluency; whereas, the second stage focuses on the participants' ability to accurately communicate in a broader range of contexts. The tasks in the CLBA listening/speaking component are as follows:

Stage I	***Listening/ Speaking Tasks***	***Stage II***	***Listening/ Speaking Tasks***
Task Type A	Follows and responds to simple greetings and instructions	**Task Type A**	Comprehends and relates video-mediated information
Task Type B	Follows and responds to questions about basic personal information	**Task Type B**	Comprehends and relates audio-mediated information
Task Type C	Takes part in a short informal conversation about personal experience	**Task Type C**	Discusses concrete information on a general topic
Task Type D	Describes the process of obtaining essential goods and services	**Task Type D**	Comprehends and synthesizes abstract ideas on a general topic

The assessment is designed so that it can be terminated at the end of any one of the four task types in stage II. A participant who begins to have difficulty with the standardized listening prompts or is unable to express complex ideas fluently is

This information is taken from the Canadian Language Benchmarks Assessment Manuals and printed with permission from the Centre for Language Training and Assessment.

BEST COPY AVAILABLE

considered to have reached his/her "threshold" – the limit of his/her proficiency. The assessment is discontinued at this point in order to prevent the experience from being uncomfortable or intimidating. However, the assessor must probe enough to ensure that the participant's highest level of proficiency has been elicited.

The current CLBA is designed to identify 8 levels of proficiency, benchmark 1 to benchmark 8. Those who achieve benchmark 8 are considered highly proficient in both their aural and oral skills.

The **CLBA Reading Assessment** is in two stages and there are four parallel forms for each stage. The range of task types for each stage is as follows:

Stage I	*Reading Tasks*	*Stage II*	*Reading Tasks*
Task Type A	Reads simple instructional texts	**Task Type A**	Reads complex instructional texts
Task Type B	Reads simple formatted texts	**Task Type B**	Reads complex formatted texts
Task Type C	Reads simple unformatted texts	**Task Type C**	Reads complex unformatted texts
Task Type D	Reads simple informational texts	**Task Type D**	Reads complex informational texts

Note: Measurement reports on the development of the CLBA are available through the Centre for Language training and Assessment, Centre for Education and Training. These reports address questions on the development and validation of the CLBA Reading Assessment.

The **CLBA Writing Assessment** is in two stages and there are four parallel forms for each stage. The range of task types for each stage is as follows:

Stage I	*Writing Tasks*	*Stage II*	*Writing Tasks*
Task Type A	Copies information	**Task Type A**	Reproduces information
Task Type B	Fills out simple forms	**Task Type B**	Fills out complex forms
Task Type C	Describes personal situations	**Task Type C**	Conveys formal messages
Task Type D	Expresses simple ideas	**Task Type D**	Expresses complex ideas

This information is taken from the Canadian Language Benchmarks Assessment Manuals and printed with permission from the Centre for Language Training and Assessment.

BEST COPY AVAILABLE

Each task represents a different genre and becomes increasingly more complex throughout the assessment. Stage I has been developed with a familiar, personal audience in mind whereas the tasks in Stage II assume a less familiar, more formal audience.

The scoring procedures for the CLBA Writing Assessment were designed to incorporate the most effective and efficient aspects of both the holistic and analytic approaches. Assessors first evaluate the overall impression made by the writing sample in respect to the task's objectives then go on to examine some of the structural and mechanical aspects of the discourse.

Based on the level of proficiency of participants in the research project, participants were not required to take Stage I of the reading and writing components of the CLBA assessment.

This information is taken from the Canadian Language Benchmarks Assessment Manuals and printed with permission from the Centre for Language Training and Assessment.

Overview of the TOEFL

The purpose of the TOEFL test is to evaluate the English proficiency of people whose native tongue is not English. The test was originally developed to measure English proficiency of international students intending to study at colleges and universities in the United States and Canada, and this continues to be its primary function. The TOEFL test is recommended for students of the eleventh grade level or above as the test content is considered too difficult for younger students.

The test is made up of four sections, which include:

- Listening
- Structure/Writing
- Reading
- Writing (Essay Rating)

The test itself is primarily computer-based as the paper-based version is being phased out. The TOEFL test utilizes two types of computer-based testing: computerized linear and computer adaptive. Two sections (Listening and Structure) are computer-adaptive and one section (Reading) is linear.

In a **linear test**, examinees are presented with questions that cover the full range of difficulty (from easy to difficult) as well as the content specifications designated by the test design. In the reading section, questions are selected without consideration of examinee performance on the previous questions.

In a **computer-adaptive test**, each examinee receives a set of questions that meet the test design and are generally appropriate for his or her performance level. The computer-adaptive test starts with questions of moderate difficulty. As examinees answer each question, the computer scores the question and uses that information, as well as responses to previous questions, to determine which question is presented next. As long as examinees respond correctly, the computer typically selects a next question of equal or greater difficulty. In contrast, if they answer a question incorrectly, the computer typically selects a question of lesser or equal difficulty.

The **Listening Section** measures the ability to understand English as it is spoken in North America. Conversational features of the language are stressed, and the skills tested include vocabulary and idiomatic expression as well as specific grammatical constructions that are frequently used in spoken English. This section includes various stimuli, such as dialogues, short conversations, academic discussions, and mini-lectures, and poses questions that test comprehension of the main ideas, the order of process, supporting ideas and inferences, as well as the ability to categorize topics/objects. This section consists of 30-50 questions and is 40-60 minutes in length.

This information is taken from the TOEFL Computer-Based TOEFL Score User Guide (1998-99 Edition) and reprinted by permission of Educational Testing Service, the copyright owner.

The **Structure Section** measures the ability to recognize language that is appropriate for standard written English. The language tested is formal, rather than conversational. The topics of the sentences are associated with general academic discourse. These are questions in which examinees must (1) complete an incomplete sentence using one of four answers provided and (2) identify one of four underlined words or phrases that would not be accepted in English. There are 20-25 questions in this section, which is 15-20 minutes long.

The **Reading Section** measures the ability to read and understand short passages similar in topic and style to academic texts used in North American colleges and universities. Test items refer to what is stated or implied in the passage, as well as to words used in the passage. This section consists of the following types of questions:

1) traditional multiple-choice questions;
2) questions that require examinees to click on a word, phrase, sentence, or paragraph to answer;
3) questions that ask examinees to "insert a sentence" where it fits best.

The Reading section includes 44-60 questions and is 70-90 minutes long. The section consists of four to five passages of 250-350 words, with 10-14 questions per passage.

The **Writing Section** measures the ability to write in English, including the ability to generate, organize, and develop ideas, to support those ideas with examples or evidence, and to compose a response to one assigned topic in standard written English. The essay rating is incorporated into the Structure/Writing scaled score and constitutes approximately 50 percent of that combined score. The rating is also reported separately on the Official Score report to help institutions better interpret examinee's Structure/Writing scores.

This information is taken from the TOEFL Computer-Based TOEFL Score User Guide (1998-99 Edition) and reprinted by permission of Educational Testing Service, the copyright owner.

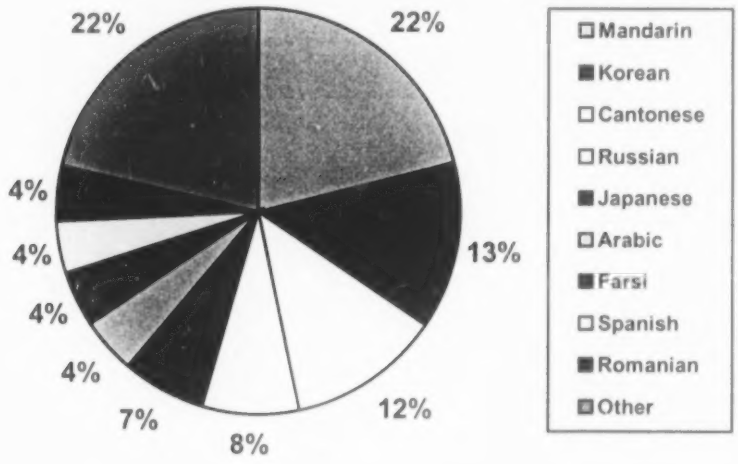
Description of the participants

From the 121 individuals who participated in the study, 90 provided valid TOEFL record scores and were included in the study. The 31 participants who were excluded from the study either had TOEFL scores that were too old for meaningful comparison, or were unable to provide a copy of their TOEFL record score. From both background data and descriptive statistical data, the 90 participants included in the study can be characterized as highly representative of the academically oriented audience who require TOEFL for university or college admission. The profile of the sample is as follows.

Country of Origin and Mother Tongue

The participants represented 26 countries of origin, with 10 countries accounting for approximately 75% of the group. The top six countries in terms of frequency of participants were: China, Korea, Hong Kong, Taiwan, Russia and Japan. Participants in the study also reported 21 different mother tongue languages, with 10 languages accounting for 80% of the sample population.

Figure 1
Distribution of Participants by Mother Tongue



Length of Residence

Length of Residence in Canada was determined from the date of entry into Canada to the date of the CLBA testing. Length of Residence varied from less than one month to nearly 16 years. The average length of residence for the population was approximately 22 months, with a median of 18 months. The mean/median distribution of LOR was characteristic of a newly arrived population.

Previous Educational/Professional Status

The participant sample represents a highly educated and professional group. On average, participants reported three years of college or university experience. Approximately 15% held graduate degrees, 30% held undergraduate degrees, 25% held college diplomas and 30% held high school diplomas. About 48% of the participants had professional careers prior to arriving in Canada. About 50% of the professions related to Health Sciences, Engineering or Physical Sciences, including such professions as: Civil Engineering, Chemical Engineering, Medicine, Nursing, and Computer Science. Other professions that were frequently reported included: Geology, Financial Planning, Law, Business and Teaching. Approximately 52% of the participants had no previous professional experience and had been full time students prior to arriving in Canada.

Present Educational/Professional Status

Approximately 63% of the participants reported no Canadian work experience. They were either seeking admission to universities and colleges or presently enrolled in universities and colleges. Of the 37% who reported Canadian work experience, the vast majority was employed in clerical/customer service, manual labour, or para-professional assistant positions. There was a clear distinction between the previous professional status of participants and their present status. Present work experience was frequently reported as: Cleaner, Gas Station Attendant, Fast Food Worker, Sales Assistant, Waitress and Security Guard.

The present educational status of the participants was divided between those who were presently enrolled directly in universities and colleges (42%), those who were seeking admission either through preparation programs or through direct application (51%) and those who were seeking professional licensing (7%). In other words, the participant group represented both those whose English Language Proficiency met the requirements of college/university admission with a TOEFL score of 560 (paper-based) or 220 (computer–based), and those whose English Language Proficiency was below the TOEFL cut score required for admission.

TOEFL score comparison for the participant sample

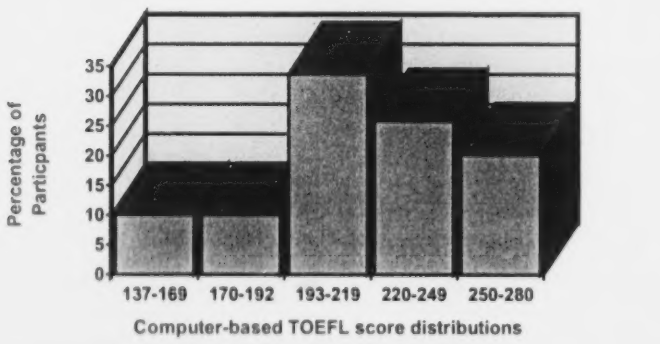
While the participant sample reflects a broad based representation of academically oriented individuals, a further comparison of the group's performance on the section scores reported on the TOEFL record was conducted. Inter-correlations for the section scores of the sample population were performed and compared to the same inter-correlations that have been reported by Educational Testing Service for TOEFL test takers between 1995-1996. These correlations demonstrate a similar item structure on the sub-scales for the two samples. These are represented in Table 1, below.

Table 1:
Inter-correlations Among Scores for ETS TOEFL Audience and Sample Population

Audience	Listening Comprehension	Structure & Written Expression	Reading Comprehension
ETS Total TOEFL Score	.86	.92	.92
Sample Population Total TOEFL Score	.79	.89	.87

The TOEFL score distributions for the sample population ranged from a low of 137 (computer) or 457 (paper) to a high of 280 (computer) or 653 (paper), with a median and mean score of 217 (computer) or 553 (paper). The sample population's TOEFL scores can be generalized as follows: 46.1% received scores sufficient for university admission at 560/220 or better, 33.7% received scores representative of advanced English preparation programs, between 530/193 and 559/219, and 20.2% received scores from 191/520 to 137/457. This information is graphically represented in the Figure 2, below.

Figure 2
Distribution of TOEFL Scores



CLBA Benchmarks Distribution for the Sample Population

The sample population's Canadian Language Benchmarks ranged largely between benchmarks 4-8. At present, the CLBA (the assessment tool for the Benchmarks) is only developed up to Benchmark 8. In order to accommodate for this limitation, we used a combination of means to identify individual performances in the three sections of the CLBA that exceeded the proficiency standards described at Benchmark 8. Our goal was to discriminate between performances that had reached their threshold at Benchmark 8 from those that were suggestive of a capacity to perform beyond Benchmark 8. This latter group was identified simply as Benchmark 8+. Decisions about 8+ status were made through a combination of assessor decision-making and participant performance on the individual tasks of the Listening and Speaking section. Cumulative scale score performances on the Writing section were used to discriminate threshold 8 performance from 8+ performance on the CLBA Writing section. Identification for 8+ on the Reading section was determined statistically, using one standard deviation above the group mean on the cumulative error count of the raw score as the cut-off. Statistical methods of selection were only available for the Reading section, as both of the other two sections rely on assessor evaluation to establish the benchmark. For the sample population in this study, we identified a total of 11.1% of the possible section performances (90 participants X 3 section scores) as 8+. Performance on the CLBA Reading section provided the largest percentage of 8+ scores, suggesting either that sample population was more proficient at receptive reading comprehension than any other skill, or that the CLBA Reading component for Benchmarks 5-8 is less difficult than the other two sections. Table 2 presents the distribution of the CLBA results for the three sections.

Table 2: CLBA Score Distribution for the Sample Population

CLBA	BM 4	BM 5	BM 6	BM 7	BM 8	BM 8 +
Listening/ Speaking	5.6% n=5	22.2% n=20	7.8% n=7	21.1% n=19	31.1% n=28	12.2% n=11
Reading	-	-	10.0% n=9	33.3% n=30	32.2% n=29	24.4% n=22
Writing	3.3%n n=3	12.2% n=11	42.2% n=38	28.9% n=26	5.5% n=5	7.8% n=7

The information in Table 2 is represented graphically in Figures 3 through 6. Figure 3 presents the overall findings of frequency for each benchmark, while Figures 4 through 6 present the distribution of Benchmarks for each section of the CLBA (Listening/Speaking, Reading, Writing). A visual comparison of the distribution of TOEFL scores (Figure 2) with the three individual sections of the CLBA suggests that there is a degree of similarity in the distributions of performances in Reading (Figure 5) and, to a lesser degree, in Writing (Figure 6), but that Speaking/Listening performances vary widely.

Figure 3
Canadian Language Benchmark Assessment Outcomes by percentage of participants

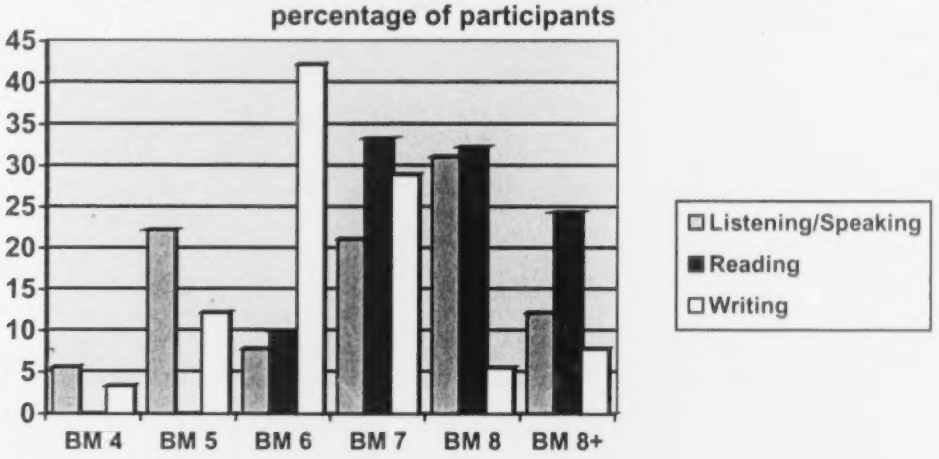


Figure 4
Distribution of CLBA Listening/Speaking

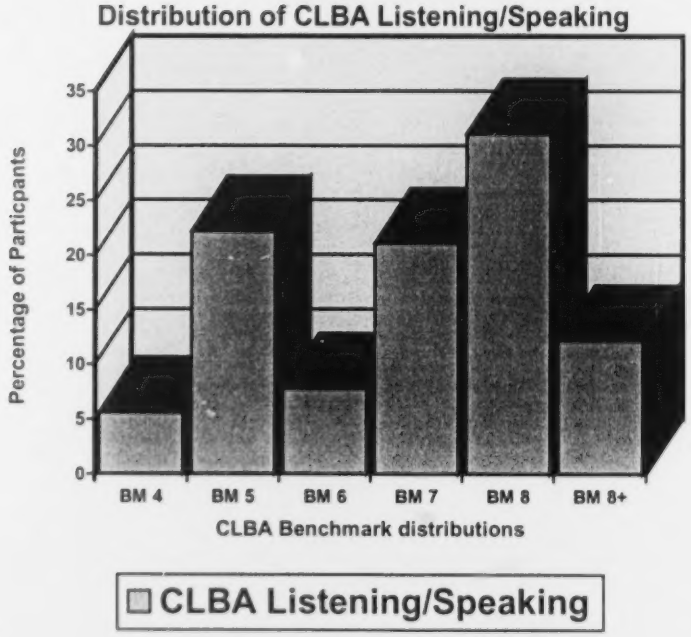


Figure 5
Distribution of CLBA Reading

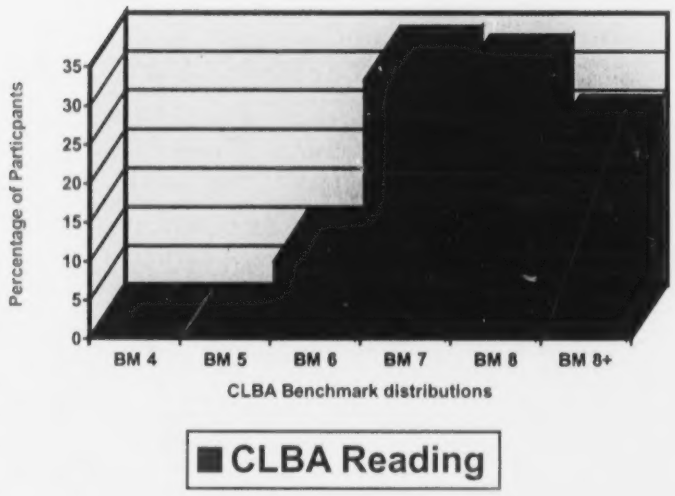
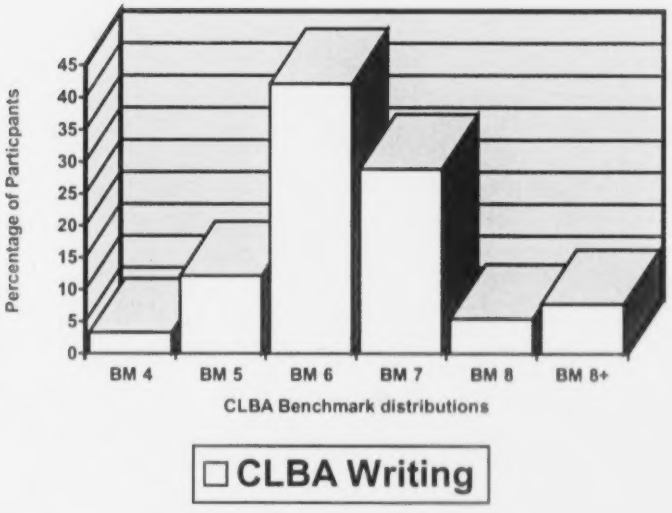


Figure 6
Distribution of CLBA Writing



At the time of this report, the CLBA has only been developed up to the end of Stage II (Benchmarks 5-8). It has been widely assumed that Stage III development (Benchmarks 9-12) would be required in order to measure the English language proficiency of academically oriented individuals who were seeking admission to universities and colleges. Our sample, which represents individuals who are seeking admission and individuals that have been admitted based on acceptable TOEFL scores, suggests a different interpretation. While there is no doubt that the CLBA will benefit from the development of Stage III (Benchmarks 9-12), Stage II (Benchmarks 5-8) is capable of assessing the English language proficiency of a sample population with a range of TOEFL scores that cluster around the university and college TOEFL cut score of 220/560.

Correlations among CLBA and TOEFL Scores

TOEFL scores officially stale-date after 2 years and it is a general practice to stale-date CLBA scores after six months. In order to increase the comparability of the two tests scores, we sought to limit the time span between to the two tests as much as possible. The majority of the participants had completed the TOEFL test prior to taking the CLBA, however, about 25% of the participants took the CLBA prior to the TOEFL. In this study, the time span between the two tests for the sample population averaged 6 months, with a median time of only 3 months. This presents a reliable comparison for the two tests, given the limited development in English language proficiency that can occur in the time frame and the balance between the orders in which the tests were taken.

Table 3 reports Pearson correlations between the TOEFL total score and the three sections of the CLBA. Further, each section score of the TOEFL was correlated with the corresponding CLBA section, measuring the similar language construct. TOEFL listening was correlated with CLBA listening/speaking. TOEFL structure/writing and TOEFL essay rating were correlated with CLBA writing. And finally, TOEFL reading was correlated with CLBA reading. The results of the correlational analysis are presented below.

Table 3: Pearson Correlations for TOEFL – CLBA Comparisons

	CLBA Listening/Speaking	CLBA Reading	CLBA Writing
TOEFL Total Score	.4183	.5102	.6221
TOEFL Listening	.5615	N/A	N/A
TOEFL Reading	N/A	.4524	N/A
TOEFL Structure/Writing	N/A	N/A	.5350
TOEFL Essay Rating	N/A	N/A	.2402

The correlation of the TOEFL total score with the three sections of the CLBA resulted in a moderately strong and statistically significant correlation (P=.000), ranging from .41 to .62. This suggests that the two tests are measuring similar language constructs, but each may provide unique information about the participant's English language proficiency. Other TOEFL comparison studies, such as the TOEFL-TOEIC comparison (Chauncey Group, 1999) also note moderate correlations (Listening .65, Reading .68, Total Score .71, n=103) and draw similar conclusions about the degree of commonality and uniqueness. High correlations in the range of .80 and above would be suggestive of convergent validity and would point to the potential to substitute one test for the other. Moderate correlation, on the other hand, argues more for the potential and value of including the two tests in the same category, for academic admission purposes.

The moderate correlation may relate to the differences in the express purposes of the two tests, or to the underlying view of language taken by the two measures. In its description of the use of TOEFL test scores (Test and Score Manual 1997) Educational Testing Service provides the following description of theTOEFL test.

> "The TOEFL test is a measure of general English proficiency. It is not a test of academic aptitude or of subject matter competence, nor is it a direct test of English speaking or writing ability. TOEFL Scores can assist in determining whether an applicant has attained sufficient proficiency in English to study at a college or university."(p. 25)

The CLBA is intended as a measure of general English language proficiency as it relates to personal communication, career/professional communication and daily life situations. It is largely aimed at assessing the language ability of adults for integration into employment related contexts.

The correlations among the TOEFL sections and the related CLBA sections also demonstrate a statistically significant and moderate set of correlations (listening/speaking .562, reading .452, writing .535) with the notable exception of the essay rating section of the TOEFL (.240), which was not significantly related. The low correlation between the TOEFL essay rating section and the CLBA writing section may be related to the inability of participants to prepare in advance for the practically oriented writing tasks of the CLBA. Essay writing on the TOEFL is a well-known phenomenon and test takers are able to practice the demand of the expository genre, before taking the test. The CLBA writing tasks were an unknown commodity to the test takers.

The correlation between the reading sections of the TOEFL and the CLBA provide a moderate comparison, at .452. Given the large number of participants who scored in the 8+ range on the CLBA (one standard deviation above the mean, based on the cumulative error for the raw scores on the CLBA reading tasks) it would seem that the TOEFL reading section is able to measure higher levels of reading difficulty than are available on the CLBA, in Stage II. Nonetheless, performance on the CLBA reading section is a moderately good predictor of TOEFL performance.

Listening/Speaking, as might be expected by the blended category, has an anticipated lower correlation to the total TOEFL score (.418), though correlates moderately with performance on the Listening section of the TOEFL (.562). An investigation of the TOEFL total scores for those who were identified as 8+ in CLBA speaking/listening showed a range of total TOEFL scores between 250 and 280. This descriptive statistic led to the consideration of the effects of Length of Residence on the measures of English language proficiency in the two tests.

Length of Residence

Length of Residence (LOR) has been commonly hypothesized as a critical factor in the development of communicative competence (Klesmer, 1990; Collier 1987, 1989). Working and communicating in an English language context has an impact on the fluency and familiarity with the general expectations for communication. For this study, we defined LOR as the number of months that an individual had resided in Canada.

To determine the effect of LOR on the various measures of English Language proficiency on the TOEFL and the CLBA, simple regression analysis was performed for LOR and each of the sections of the two tests, including the TOEFL total score. None of the dependent variables with the exception of the CLBA listening/speaking section were significantly predicted by Length of Residency. LOR was held to be highly predictive of performance on the CLBA listening/speaking section (Significance of F=.02). This finding suggests that the CLBA has a valuable contribution to make to the process of determining the proficiency of individuals in communicating in real life contexts. The fact that none of the other sections on the two tests proved significant as predictors of performance strengthens the potential contribution that the CLBA listening/speaking section can make to the general assessment of English language proficiency. Table 4 summarizes the results of the Regression analysis.

Table 4: Regression Analysis for Length of Residence

	F Value	Significance of F
Total TOEFL	.10938	.7416
TOEFL Listening	.92557	.3387
TOEFL Reading	.00018	.9892
TOEFL Structure/Writing	.0074	.9784
TOEFL Essay Rating	1.40394	.2393
CLBA Listening/Speaking	5.59819	.0202*
CLBA Reading	1.39022	.24515
CLBA Writing	2.70975	.1033

* p< .05

Equating TOEFL Performance with CLBA Performance

Equating test scores on different test instruments requires a large and purposefully stratified sample. However, the findings in this study were sufficiently robust to establish a heuristic comparison chart. In order to provide a degree of consistency with ETS procedures for concordance tables, a scaled score equation algorithm used by ETS was applied to the TOEFL total scores. This produced scaled equation scores for the three sections of the CLBA, based uniquely on the interpolation of real score comparisons, generated from the study. Table 5 presents the heuristic concordance. The table is divided into shaded ranges. Each range represents the established range comparison by ETS for its computer-based and paper-based scores. Ranges are calculated to coincide with the Standard Error of Measurement (SEM) for TOEFL performance. ETS defines SEM as follows:

> "The standard error of measurement (SEM) is an estimate of the probable extent of error inherent in a test score due to the imprecision of the measurement process". (1997:30)

The scaled CLBA Benchmarks, which are reported in Table 5, are decimalized, to represent the continuum of scaled scores. The CLBA is not a decimalized scale of benchmarks. Therefore, a conservative approach of upward rounding all decimals is suggested in order to account for both the heuristic nature of the scale and the possible negative effects of inter-rater reliability, as it applies to the CLBA (Cohen, 1999). Only scores between 137 and 280 were reported, since these represent the actual scores for the participant sample.

A measure of reliability was gained by avoiding extrapolation (the process of statistically inferring score equivalencies beyond existing data). Nonetheless, it is important to note that the resulting heuristic concordance table is a preliminary attempt to establish the feasibility of concordance between the CLBA and the TOEFL test. Interpolation of the scores captures scaled equation scores for Benchmarks 3-9 in the three skill areas.

Table 5:
Algorithmic Concordance Table – Total Scaled Scores Comparison[1]

TOEFL Computer-based Total	TOEFL Paper-based Total	CLBA Listening/ Speaking	CLBA Reading	CLBA Writing
137	457	3.52	5.88	4.04
140	460	3.64	5.94	4.13
143	463	3.77	6.00	4.21
150	470	4.05	6.14	4.42
157	480	4.34	6.29	4.63
160	483	4.46	6.35	4.72
167	493	4.74	6.49	4.93
170	497	4.86	6.55	5.01
177	503	5.15	6.69	5.22
180	507	5.27	6.75	5.31
183	513	5.39	6.81	5.40
187	517	5.56	6.89	5.52
190	520	5.68	6.95	5.61
193	523	5.80	7.01	5.70
197	527	5.96	7.09	5.82
200	533	6.09	7.15	5.90
203	537	6.21	7.21	5.99
207	540	6.37	7.29	6.11
210	547	6.49	7.35	6.20
213	550	6.62	7.41	6.29
217	553	6.78	7.49	6.41
220	560	6.90	7.55	6.55
223	563	7.02	7.61	6.59
227	567	7.19	7.69	6.70
230	570	7.31	7.75	6.79
233	577	7.43	7.81	6.88
240	587	7.72	7.95	7.09
243	593	7.84	8.01	7.18
247	597	8.00	8.09	7.30
250	600	8.12	8.15	7.39
253	607	8.25	8.21	7.48
257	613	8.41	8.29	7.59
260	620	8.53	8.35	7.68
263	623	8.65	8.41	7.77
267	630	8.82	8.49	7.89
273	637	9.06	8.61	8.19
277	647	9.22	8.69	8.19
280	653	9.35	8.75	8.28

[1] The summary table was compiled using an algorithmic procedure from: *Summary and Derivation of Equation Methods used at ETS*, W.H. Angoff, 1982. In: **Test Equation** Holland & Rubin (eds), Academic Press.

BEST COPY AVAILABLE

Technical Note On Equating TOEFL & CLBA Performances

Statistical Equating is the process of developing a conversion from the system of units of one form of a test to the system of units of another form so that scores derived from the two forms after conversion will be equivalent and interchangeable. Linear equating, which we employ here, is one of the general methods of equating. It is based on the following definition: Two scores, one on Form x and the other on Form y - again, where x and y are equally reliable and parallel measures - may be considered equivalent if their respective standard score deviations in any given group are equal:

$$(y-M_y)/S_y = (x-M_x)/S_x$$

where M_y, M_x are means for form y, and x respectively,
S_y, S_x are standard deviations for form y, and x respectively.

When these terms are rearranged, we have:

$$y=M_y + (x-M_x)*S_y/S_x$$

In our case, we treat x as: the TOEFL total score, and y as different CLBA scores. From our research sample, we have the following descriptive statistics:

Variable	Mean	Std Dev	Minimum	Maximum
TOEFL Total	216.04	33.39	137.00	280.00
CLBA Lis/Speak	6.74	1.36	4.00	8.00
CLBA Reading	7.47	.67	6.00	8.00
CLBA Writing	6.38	.99	4.00	8.00

So our scale scores for each component of CLBA are:

CLBA: Listening = (TOEFL total score - 216.04)*1.36/33.39+6.74.

CLBA: Reading = (TOEFL total score - 216.04)*.67/33.39+7.47.

CLBA: Writing = (TOEFL total score - 216.04)*.99/33.39+6.38.

Reference: Test Equating, Edited by: Paul W. Holland & Donald B. Rubin. Page 55-69, Academic Press, 1982.

Conclusions

The findings in this report are of value to a variety of areas concerned with English Language Proficiency standards for academic and professional purposes. Foremost among these is the value to universities and colleges in establishing equivalence between the CLBA and the TOEFL for admission purposes. The heuristic concordance table provides the basis for an informed discussion of equivalent standards. While the table represents a preliminary concordance, it demonstrates a reliable comparison within the limitations of this study.

The findings may also be of use in setting curricular thresholds for Academic English preparation programs in universities and colleges. Within many institutions, successful completion of an internal EAP program meets the English language proficiency requirements for admission to degree granting programs. By structuring Academic English Preparation programs along the expected standards of performance as described in the Canadian Language Benchmarks, it may be possible to establish recognition of internal institutional standards of equivalence across universities and colleges, thereby increasing the portability of previous study. This would further improve educational access to degree programs for second language speakers of English.

Professional associations and their regulatory bodies would also benefit from the concordance of the CLBA with their existing measures for establishing professional standards for English language proficiency. The CLBA assesses a range of personal, professional and daily life communication contexts. The Listening/Speaking section of the CLBA further offers a reliable and nationally available oral proficiency interview, and therefore provides a direct assessment of the communicative ability of potential applicants in face-to-face communication.

Lastly, the methodology and statistical procedures used in this study provide a replicable basis for future studies and their meta-analysis. They establish the comparability of the participant group to a larger sample of TOEFL takers, enhancing the generalizability of the findings. They also adhere to statistical procedures that are commonly used in establishing concordance equivalencies, allowing for the future comparison of other tests.

Recommendations

1. Undertake a broader scaled study to verify the heuristic table of concordance between the TOEFL and the CLBA. Increasing the sample size and adding the CLBA Stage III (Benchmarks 9-12) once developed, would add to the reliability of the initial concordance reported in this study.

2. Promote the use of the CLBA for admission purposes in Canadian universities, colleges and professional associations.

3. Undertake similar studies with professional associations in order to update English language proficiency standards for professional standing.

4. Compare the CLBA to other accepted equivalencies for university and college admission (e.g. three years of full time study in a Canadian institution, or five years of residency, etc.).

5. Include the findings of this research project in the formative development of the CLBA Stage III (Benchmarks 9-12).

Researcher Information

David L.E Watt
David Watt is an associate professor and co-ordinator of the MEd. TESL program in the Faculty of Education at the University of Calgary. He is a nationally certified CLBA assessor and has published research in the areas of: ESL dropout, English Language Proficiency and educational adjustment, and educational policy in English as a Second Language.

Deidre M. Lake
Deidre Lake is the ESL/LINC program manager for the Calgary Mennonite Centre for Newcomers, a large, community based ESL provider. She is a nationally certified CLBA assessor, who has conducted over 3,000 assessments using the CLBA. She has published classroom assessment tasks related to the CLBA and works as a freelance writer/researcher on assessment issues in language and literacy.

For more information about the Canadian Language Benchmarks, contact:

Centre for Canadian Language Benchmarks
200 Elgin Street, Suite 803
Ottawa, ON. K2P 1L5

Telephone: 613-230-7729
Fax: 613-230-9305
E-mail: info@language.ca

Web site:http://www.language.ca

For more information on the TOEFL test, contact:

Test of English as a Foreign Language
Educational Testing Service
P.O. Box 6155
Princeton, NJ. 08541-6155

Telephone: 609-771-7100
E-mail: toefl@ets.org

Web site:http://www.toefl.org

For information about the research project, contact:

David L.E. Watt, Associate Professor
University of Calgary
Education Tower 718
2500 University Dr. NW
Calgary, AB. T2N 1N4

Telephone: 403-220-7353
Fax: 403-245-4110
E-mail: dwatt@ucalgary.ca